

Themed Issue: Process Analytical Technology

Guest Editor - Ajaz Hussain

Process Analytical Technology Case Study, Part III: Calibration Monitoring and Transfer

Submitted: July 27, 2004; Accepted: December 13, 2004; Published: October 6, 2005

Robert P. Cogdill,¹ Carl A. Anderson,¹ and James K. Drennen III¹

¹Duquesne University Center for Pharmaceutical Technology

ABSTRACT

This is the third of a series of articles detailing the development of near-infrared spectroscopy methods for solid dosage form analysis. Experiments were conducted at the Duquesne University Center for Pharmaceutical Technology to develop a system for continuous calibration monitoring and formulate an appropriate strategy for calibration transfer. Indicators of high-flux noise (noise factor level) and wavelength uncertainty were developed. These measurements, in combination with Hotelling's T^2 and Q residual, are used to continuously monitor instrument performance and model relevance. Four calibration transfer techniques were compared. Three established techniques, finite impulse response filtering, generalized least squares weighting, and piecewise direct standardization were evaluated. A fourth technique, baseline subtraction, was the most effective for calibration transfer. Using as few as 15 transfer samples, predictive capability of the analytical method was maintained across multiple instruments and major instrument maintenance.

KEYWORDS: process analytical technology (PAT), near-infrared spectroscopy (NIR), tablet analysis, pharmaceutical analysis, calibration transfer.

INTRODUCTION

This is the third of a series of articles detailing the development of near-infrared (NIR) methods for solid dosage form analysis. The first article described early feasibility studies to verify the performance of a new online tablet analyzer and provided vital data for calibration development.¹ The second detailed the development and validation of quantitative NIR calibrations for tablet active pharmaceutical ingredient (API) content and hardness.² This article examines the details of implementing and managing NIR calibrations to ensure the continuity of predictive performance. These

details are summarized in terms of calibration monitoring and calibration transfer.

Published investigations of process analytical technology (PAT)³ applications often focus on selection of technology, and the development and validation of predictive calibrations. While these are essential aspects of a PAT application, a suitable infrastructure must be in place to support the deployment, operation, and maintenance of the PAT method in a real-time manufacturing environment. The development of this infrastructure comprises the third phase of PAT method development.

CALIBRATION MONITORING

For every NIR tablet analysis performed, there are 4 possible outcomes that depend on the state of the tablet and the state of the measurement system:

1. The tablet is within specification, and the measurement is valid.
2. The tablet is within specification, and the measurement is not valid.
3. The tablet is out of specification, and the measurement is valid.
4. The tablet is out of specification, and the measurement is not valid.

Any given measurement provides information about the product and the operational capabilities of the measurement system. The states of the measurement system and product determine whether the analytical method, the product, or both require investigation and corrective action. There is finite probability that an analytical result will be rendered invalid as a result of sampling errors, instrumental drift or failure, or product changes. If the prediction failure is the result of sampling error, corrective action may be as simple as repeating the tablet analysis. For measurement failures resulting from instrumental and product variation, more substantive action is required in the form of calibration transfer or update; both will be discussed in the following sections. In any event, it is imperative that a robust protocol be in place to monitor the state of the system, and, if necessary, to determine the appropriate corrective action. The purpose of calibration monitoring is to ensure that conditions (2) and (4) are detected.

Corresponding Author: James K. Drennen III, Duquesne University Center for Pharmaceutical Technology, Pittsburgh, PA 15282. Tel: (412) 396-5520; Fax: (412) 396-4660. Email: drennen@duq.edu

A significant advantage of spectroscopic PATs (eg, NIR, Raman, or NIR imaging) is the wealth of ancillary information captured with each measurement. The shape of an NIR spectrum, for example, is determined not only by the chemical and physical properties of the sample but also by the condition of the instrument and the nature of the interaction between the sample and the instrument. It was demonstrated earlier in this series^{1,2} that, with the application of appropriate chemometric techniques, it is often possible to generate accurate predictions for specific sample quality attributes, while suppressing the spectral effects of other concomitant variations. In much the same way, multivariate data analysis and signal processing techniques can be applied to monitor the state of the NIR measurement system. These techniques represent an application of multivariate statistical process control (MSPC).^{4,5} The goal of a calibration monitoring protocol is to determine whether the instrument is functioning properly and whether the calibration models are appropriate for the current sample.

CALIBRATION TRANSFER

Calibration transfer is generally regarded as a mathematical procedure whereby a model or data are transformed so they are compatible with multiple analytical instruments.⁶ For example, an NIR calibration model developed using a laboratory-based “master” spectrometer may require modification to be transferred for use with an online “slave” process spectrometer. However, calibration transfer covers a wide range of operations for the perpetuation and propagation of PAT methods. Furthermore, while they will be illustrated herein as spectroscopic analysis tools, calibration transfer techniques are generally applicable to multivariate data used for predictive modeling. Calibration transfer operations can be classified as either data transformation or model update techniques.

Calibration data transformation techniques are used either to format data for compilation from multiple sources (or events), or to adjust a set of model parameters to maintain or extend their applicability. Model adjustments for calibration transfer include linear “slope and bias” adjustments on predictions.⁷ The choice of whether the calibration model or the calibration data should be modified for calibration transfer has little bearing on predictive performance.⁸ The modification, or standardization, of spectral data are advantageous to model transformation since it facilitates cross-platform data sharing and archiving. This approach has the benefit of maintaining a single data (and model) basis. For these reasons, only transformation techniques pertaining to the standardization of spectral data are considered in this work.

A range of data transformations have been developed that can be used for calibration transfer. These techniques can

be arranged on a continuum according to complexity. The simplest data transformations are signal processing techniques such as normalization, parametric baseline adjustment (eg, detrending), bandpass filtering, and derivatization.⁶ Signal processing, or preprocessing, techniques are an important part of robust⁹ chemometric model development for feature enhancement and background suppression. They play a preventive role in calibration transfer since many instrumental and sampling effects appear as low-frequency effects such as additive and multiplicative baseline shift.¹ Many calibration transfer problems, however, require more complex, empirical data transformation. Empirical data transformation functions are applied prior to spectral preprocessing operations within the analytical data flow (Figure 1). Two types of empirical calibration data transformation will be discussed: direct orthogonalization and regression-based transformation.

Similar to signal processing techniques, direct orthogonalization (DO) algorithms may be used as preprocessing during calibration development.^{7,10-13} These algorithms, including DO,¹³ orthogonal signal correction (OSC),^{7,10-12} prediction-augmented classical least squares (PACLS),¹⁴ generalized least squares (GLS),¹⁵ and their respective variants, work by subtracting or suppressing undesirable spectral variation. For example, to transfer a calibration between 2 NIR analyzers, a linear model of the spectral differences between the instruments can be created using principal component analysis (PCA):

$$\hat{X} = X(I - P_d P_d^T) \quad (1)$$

where \hat{X} = $n \times p$ matrix of spectra, orthogonalized to the principal axes of the difference spectra. X = $n \times p$ matrix of sample or calibration spectra to be corrected. P_d = Loading vectors for the PCA model, d , of the difference spectra. I = $p \times p$ identity matrix.

Subsequently, sample spectra are corrected by subtracting the significant principal components of the difference spectra. While DO techniques have been shown to be effective for calibration transfer,¹¹ and in some cases can be derived using artificial reference standards,¹⁴ they generally require recalculation of the calibration model for each application. Furthermore, if the spectral difference between instruments is significant and/or covers a large portion of the sample spectrum, very little analytical signal may remain after orthogonalization, which may reduce predictive capability.

Regression-based transfer algorithms use a data set of paired spectra to estimate a transfer function between the signals of multiple analyzers. Rather than subtracting or suppressing instrumental differences, regression-based transfer algorithms seek to “warp” spectra to a common basis. Hence, under ideal conditions, analytical information content is not

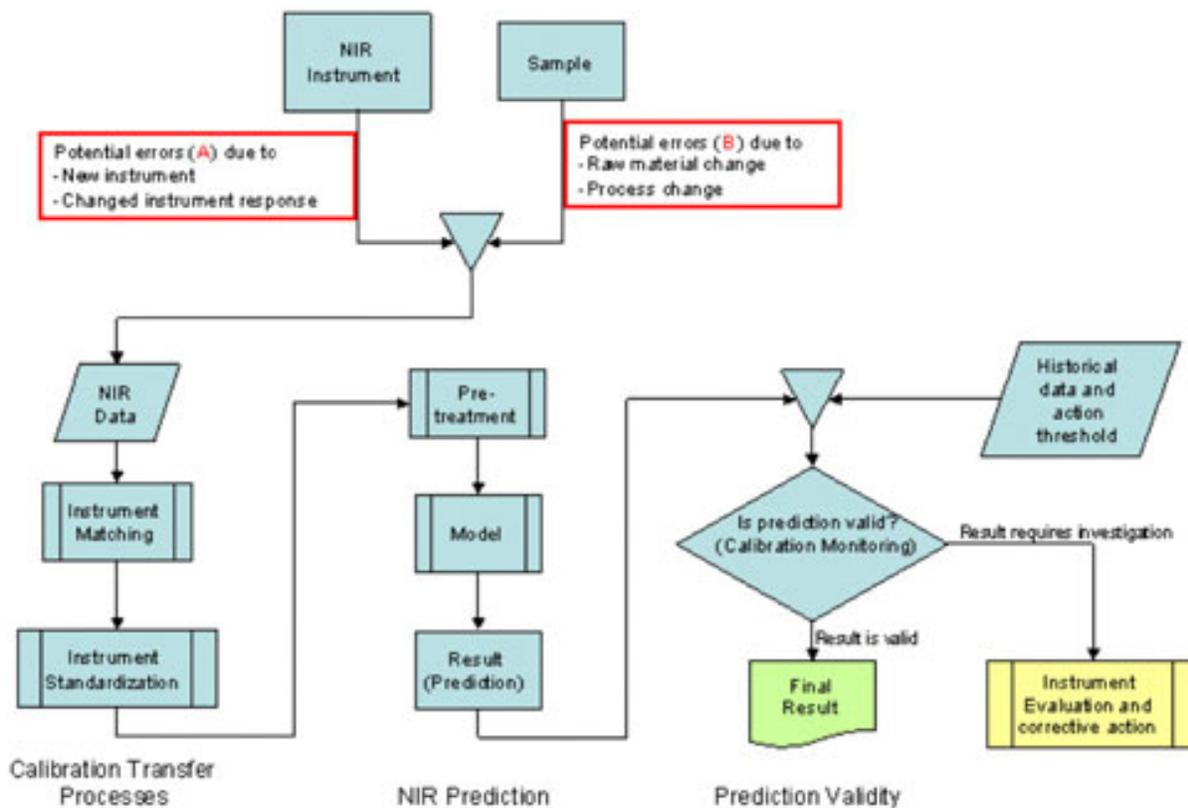


Figure 1. Data flow diagram for a PAT application of NIR spectroscopy.

compromised by the transfer. In a simple application, linear regression might be used to adjust for a gain and bias difference between 2 single-channel sensors. The patented CLONE¹⁶⁻¹⁹ algorithm, which employs quadratic regression between the paired channels of master and slave spectrometers, has been used extensively for calibration transfer and standardization among grating monochromator NIR analyzers. Direct standardization (DS)^{7,20} and piecewise direct standardization (PDS)^{7,20-22} estimate a transfer function to predict the spectral response of each wavelength of a master instrument using either the entire spectrum, or discrete spectral bands, from a slave instrument, respectively. These algorithms are capable of compensating for linear differences in sensitivity, as well as wavelength and band-pass. PDS has been used for calibration transfer between different types of spectrometers,⁷ as well as for correction of sample temperature effects.²¹ For even greater flexibility, nonlinear versions of PDS have been developed using artificial neural network (ANN) regression.⁷

For all empirical calibration transfers, great care must be taken to select the appropriate algorithm and to design the calibration transfer training data set accordingly. Regression-based calibration transfer can be as complex as the quantitative calibration development process, with increased risk of over-fit. Unlike signal processing techniques, which are implemented as an integral part of the calibration develop-

ment process, empirical calibration transfer techniques are employed in response to deviations in the NIR spectra.

Empirical calibration transfer may remedy the effects of a planned event, such as instrument maintenance (eg, lamp change), or calibration transfer to a new instrument. Alternatively, empirical calibration transfer may be required for re-standardization of the instrument following an unexpected change in instrument response. In the event that an unplanned spectral deviation is detected by the calibration monitoring protocol, and if a root cause analysis determines that the problem is the result of an instrumental change, a new calibration transfer function must be calculated by analyzing an instrument standardization sample set. In cases where no redundant master instrument is available, a “rescue set” of standardization samples and spectra should be maintained. Ideally, artificial reference standards (which are resistant to degradation) would be used for estimation of the data transformation function. This is rarely an option in practice, since it is difficult to replicate the instrument-sample interaction for process analyzers using artificial reference standards. Many researchers have observed that empirical NIR calibration transfer requires the use of real samples.^{16,17} Thus, stability tests should be performed to determine a suitable storage time for instrument standardization samples. Furthermore, a standardization sample replacement procedure should be in place prior to deployment of the PAT.

CALIBRATION UPDATE

In the event that an unplanned spectral deviation is detected by the calibration monitoring protocol and root cause analysis does not find instrumental error (state No. 2, as above), and if parallel testing finds the product to be within specifications, calibration update is required to restore analytical performance.^{19,23} Calibration or model update entails the recalculation of a predictive model with the inclusion (or exclusion) of a particular set of samples. Examples of situations that might prompt calibration update are changes in API or excipient source, or the effect of routine wear of the processing equipment, both of which may alter the NIR process signature.^{1,3} Under ideal conditions, all possible within-specification product variations are modeled during calibration development. Since this is rarely practical, quantitative NIR calibrations typically model a subset of potential variations.² Thus, the need for calibration update does not imply that the analytical method has failed; rather, new variance has been encountered that would ideally have been included in the model in the first place.

Since every scenario that might prompt the update of a PAT calibration cannot be anticipated, the nature of future calibration updates cannot be rigorously defined. However, conditions under which calibration update should be performed must be well defined. Standard operating procedures (SOPs) should be in place to guide the development and validation of an updated calibration. To date, the application presented here has not required calibration update. Therefore, calibration update will not be discussed in the following experimental sections of this paper.

OBJECTIVES

The objectives of the work reported here were the following:

1. Develop a system for continuous calibration monitoring.
2. Formulate an appropriate strategy for calibration data transformation to support instrument maintenance and interinstrument transfer.
3. Determine the required number, and allowable storage time, of instrument standardization "rescue" samples.

MATERIALS AND METHODS

NIR Instrumentation

Two Luminar 3070 (Brimrose Corp, Baltimore, MD) double-beam scanning acousto-optic tunable filter (AOTF) spectrometers were used for this work. The first spectrometer, or "master instrument," was used during the earlier phases of method development, reported in the first 2 articles of this series.^{1,2} The master instrument will be used online for

continuous monitoring of solid dosage tablets. The second spectrometer, or "slave instrument," will be used as an off-line, or backup, analyzer. Both analyzers used the automatic tablet positioning system described earlier in this series.¹ While the master and slave instruments are identical in configuration, some differences in spectral response were observed (Figure 2). For the slave instrument to serve as an effective backup instrument, it must be capable of sharing a model with the master instrument (without compromising predictive capability). Thus, a data transformation was needed to facilitate interinstrument calibration transfer.

Data Collection

Three sets of tablet samples were used for the calibration transfer investigations. The first set, or "training set," consisted of 254 tablets from multiple production-scale batches. Chemical and physical reference data were not available for the first sample set, which was used exclusively as a calibration transfer training data set. The second and third sample sets consisted of the VAL2 and VAL3 sample sets from the second article in this series.² The VAL2 data set consisted of production-scale tablets, while VAL3 consisted of laboratory-scale tablets with wide variation in API content and hardness. The VAL2 and VAL3 samples were used to test the performance of the calibration transfer algorithms tested. API content reference data were available for the VAL2 and VAL3 sample sets and were summarized earlier in this series.² Spectra of the tablets from all 3 sets were scanned using both the master and slave instruments.

A stability study was performed to determine the length of time tablets could be stored for use as calibration rescue samples. This fourth, or "stability," sample set consisted of

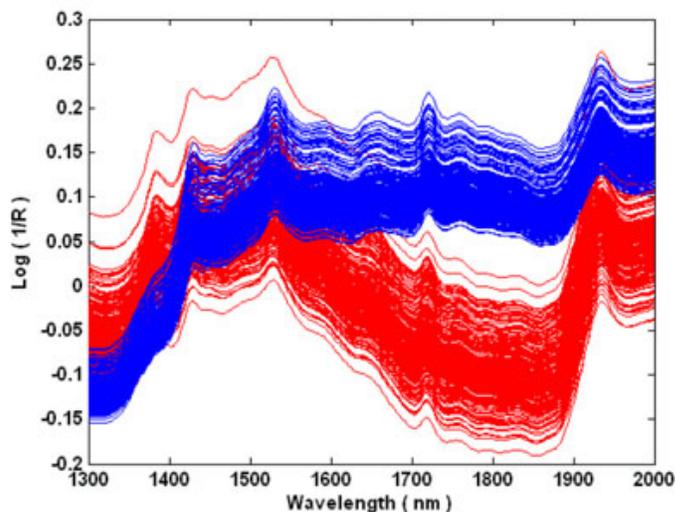


Figure 2. Raw calibration transfer training spectra for the master (red) and slave (blue) AOTF spectrometers. The same samples were measured using both instruments.

9 production-scale tablets, which were scanned repeatedly using both the master and slave analyzers over a period of 40 days. Degradation or change in NIR spectral signature was detected by application of the continuous calibration monitoring procedure.

Standard wavelength reference material SRM-1920a was used for wavelength uncertainty characterization.

The API content and hardness calibrations, and data sets,² were employed for prediction and calibration monitoring.

All data analyses were performed using Matlab 6.5 (The Mathworks, Natick, MA), and the PLS Toolbox 3.0 (Eigenvector Research Inc, Manson, WA). All studies were performed by the Duquesne University Center for Pharmaceutical Technology (DCPT).

Calibration Transfer Algorithms

Three established calibration transfer techniques were tested for their ability to maintain the predictive performance of a calibration following transfer between master and slave spectrometers. One signal processing technique, finite impulse response filtering (FIR),²⁴ and 2 empirical techniques, GLS⁵ (orthogonalization) and PDS²⁰ (regression), were investigated. A fourth technique, direct baseline subtraction,²² was evaluated.

All of the established calibration transfer methods have at least 1 parameter that must be specified. These methods were optimized using random cross-validation to specify the associated parameter values. During cross-validation, the training sets were repeatedly separated randomly (pairwise) into training and test subsets. For each iteration of FIR optimization, the API content calibration data set and the test subsets were filtered (via FIR), a temporary partial least-squares (PLS) sub-calibration was derived, and the test subsets were predicted. The mean spectrum of the API content calibration set was used as the FIR reference. The optimal FIR window width and the choice of additional preprocessing operations (eg, derivatives) were selected as the combination for which the predictions of the master and slave test subsets were most in agreement. Agreement between master and slave predictions was quantified as root mean square error of transfer (RMSET):

$$RMSET = \sqrt{\text{mean}((PRED_{mstr.} - PRED_{slave})^2)} \quad (2)$$

Once the optimal FIR parameter combination was selected, a new FIR + PLS calibration was calculated, and the FIR-transformed VAL2 and VAL3 data sets were predicted.

The GLS optimization was similar to that for FIR as it also involved recalculation of the calibration model after GLS weighting. During cross-validation, GLS submodels were

calculated using the difference between the master and slave training subsets. For GLS, the scaling parameter, a , and additional preprocessing steps were optimized. Also, a separate study (not shown) was performed to determine whether GLS weighting should be performed before or after other preprocessing operations. For this application, GLS weighting was found to perform best when performed using raw difference spectra (prior to other preprocessing).

Since PDS is independent of the calibration, only the master and slave training sets were required for cross-validation. The calibration transfer algorithms were ranked according to root mean square error of prediction (RMSEP) after optimization. RMSET was used during optimization (rather than RMSEP), since it is expected that in a "rescue" calibration transfer situation, chemical or physical reference data may not be available for the transfer sample set. Thus, if it is necessary to reoptimize the transfer method, predicted values from a master instrument would be used as reference. Furthermore, the use of any of these techniques to "improve" calibration performance is a separate issue from calibration transfer and should be considered during calibration development/update. RMSEP was used for final method selection because the effect of any transfer method on predictive performance must be considered.

RESULTS AND DISCUSSION

Calibration Monitoring

The goal of a calibration monitoring protocol is to determine whether the instrument is functioning properly and whether the calibration models are appropriate for the current sample. This determination is performed with every measurement, in real-time. The calibration monitoring protocol developed for this work is based on 4 measurements of the data: high-flux noise, wavelength shift uncertainty, lack of model fit, and sample-to-model distance.

Historically, the NIR instrument performance evaluation component of calibration monitoring has been performed off-line, or, at best, between sample scans. This common procedure is impractical for real-time use. Some instrument performance parameters can be monitored continuously by analyzing certain features of the sample spectra. For this work, tests were developed to measure high-flux noise and wavelength shift uncertainty using individual sample spectra.

The high-flux noise test is based on the analysis of the highest-frequency component of each spectrum using interpoint correlation (IPC). The scanning NIR spectrometers used for this work sample the wavelength axis at a higher frequency than the instrument bandpass, therefore high correlation is expected between absorbance intensity measurements at adjacent wavelengths. Thus, since high-flux detector noise is also manifested at a frequency beyond the instrument

bandpass, the level of IPC will vary significantly with the level of high-flux noise (Figure 3). IPC will not be significantly affected by changes in spectral shape owing to sample composition. IPC is calculated using the Pearson correlation function (r) between adjacent data points:

$$r(X_{\lambda}, X_{\lambda+1}) = \frac{n(\sum(X_{\lambda} X_{\lambda+1}) - (\sum X_{\lambda})(\sum X_{\lambda+1}))}{\sqrt{n(\sum X_{\lambda}^2) - (\sum X_{\lambda})^2} \sqrt{n(\sum X_{\lambda+1}^2) - (\sum X_{\lambda+1})^2}}, \quad (3)$$

where X_{λ} indicates odd-numbered spectral data points, and $X_{\lambda+1}$ indicates even-numbered spectral data points.

Using this formula, a noise factor level (NFL) is estimated as a function of the following:

$$NFL = f(1 - r(X_{\lambda}, X_{\lambda+1})) \quad (4)$$

With this framework in place, a function relating estimated to actual high-flux noise can be derived by using Monte-Carlo simulation (Figure 4). The NFL function does not pass through the origin because the relationship was estimated by adding artificial high-flux noise to NIR spectra, which already had some level of high-flux noise present. Furthermore, the presence of features in an NIR spectrum limits the maximum IPC, which, in turn, prevents the NFL from reaching zero.

Two methods might be used to specify control limits for the NFL. One method would be to use Monte Carlo simulation to determine a critical NFL limit, where calibration performance is significantly degraded, as was implied during validation of the API content calibration.² The second method would be to estimate limits from historical calibration development data (Figure 5). Since the calibration was shown

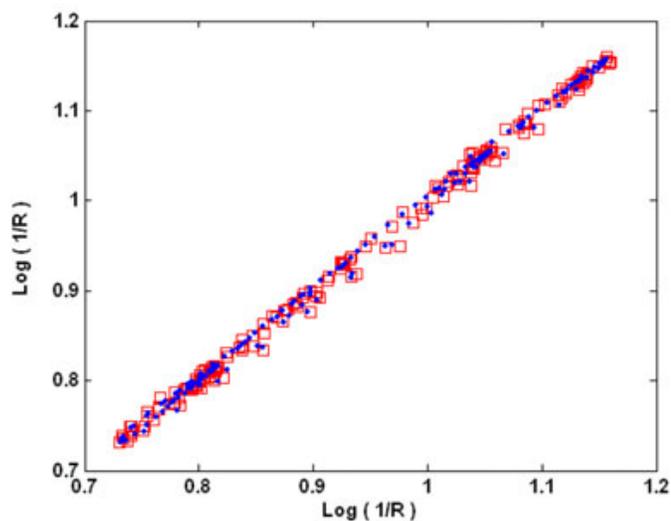


Figure 3. Example of IPC for typical (blue) and noisy (red) NIR spectra.

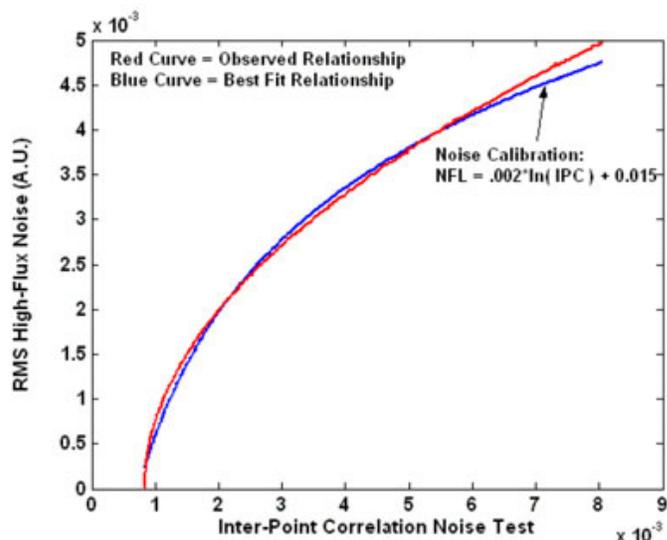


Figure 4. Estimation of an NFL transfer function relating simulated high-flux spectrometer noise to IPC.

to be robust to high flux noise levels far exceeding the historically observed NFL (Figure 5), relatively wide control limits could be justified. Regardless of the control limits, any monotonically increasing trend in NFL should be treated as an indication of pending instrument failure and should prompt preemptive investigation or maintenance. While the IPC noise test was effective for this application, other instruments (especially those with a wider spectral sampling interval) may require some alternate form of real-time noise test. One possibility might be to analyze sequential repetitive scans of each sample (without repositioning).

NIR absorbance bands may shift as the result of chemical or physical changes in the sample (or temperature variation)

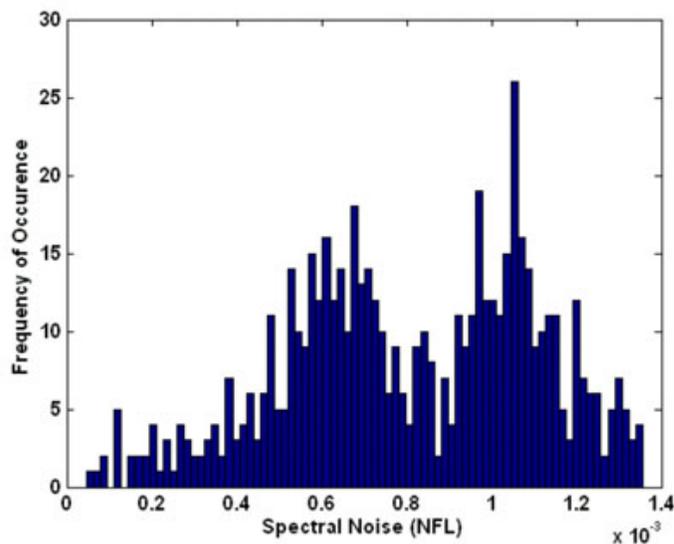


Figure 5. Histogram of historical NFL scores for the API content calibration spectra.

and instrumental error. The robustness testing performed earlier in this series² illustrated the deleterious effect of wavelength inaccuracy on the predictive capability of NIR calibrations, regardless of the source of error. A sample-based wavelength uncertainty test is performed by measuring the locations of NIR absorbance bands in the sample spectra. The centerlines of major absorbance bands are located as zero-crossings in the first or third derivative of each sample spectrum (Figure 6). Simulations were performed (not shown) to determine the accuracy of the wavelength shift test using 7 absorbance bands, centered at 1364, 1383, 1406, 1427, 1739, 1760, and 1932 nm. Test spectra were randomly offset and stretched along the wavelength axis, and the location of absorbance band centers was monitored. In all cases it was found that the added absorbance band offset was measured with less than 0.03 nm standard error.

Historical wavelength uncertainty test results (for the absorbance band centered near 1364 nm) are shown for the calibration data set in Figure 7. Wavelength shift error is distributed about a mean standard deviation of 0.22 nm. The average wavelength shift correlation coefficient for the 7 bands tested was less than 0.6. It was also observed that, for this instrumentation, absorbance band shift is not well correlated across the entire spectrum. This finding would suggest that the observed wavelength shift is either due to chemical or physical changes in the samples, or that the AOTF instrumentation exhibits complex, time-varying wavelength shift error characteristics. The wavelength shift test was then applied to multiple spectra (from the same instrument) of a standard wavelength reference material (SRM-1920a). These spectra exhibited nonconstant wave-

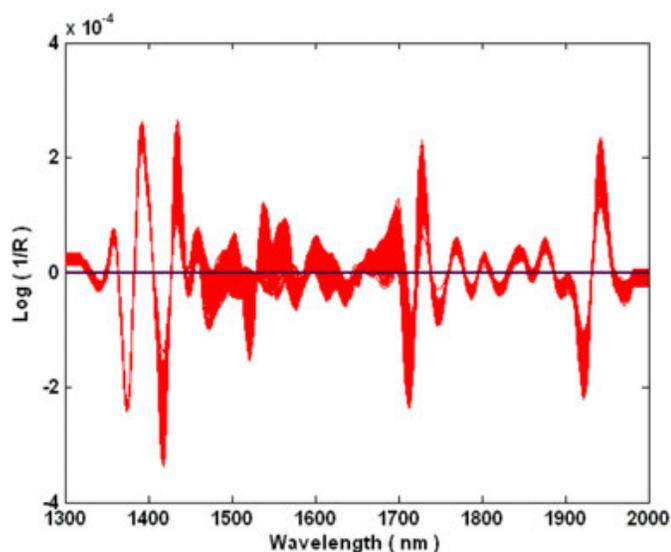


Figure 6. Third-derivative transformation of the API content calibration spectra. Major zero-crossings (used for wavelength uncertainty testing) are located near 1364, 1383, 1406, 1427, 1739, 1760, and 1932 nm.

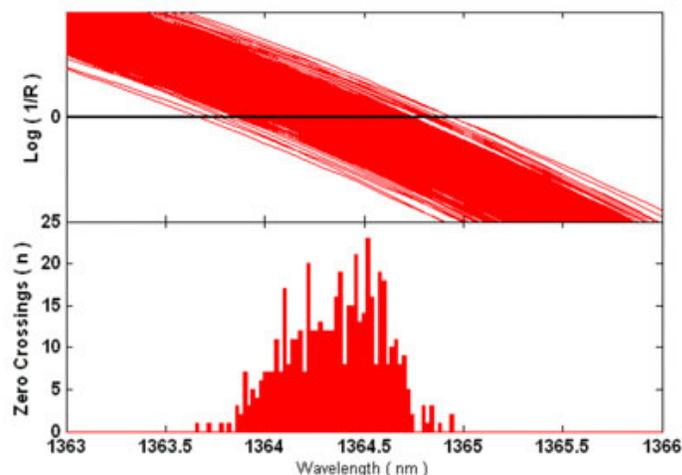


Figure 7. Demonstration of AOTF wavelength uncertainty at 1364 nm, measured using continuous calibration monitoring. The upper figure shows the third derivative traces for the API content calibration spectra from 1363 to 1366 nm. The lower figure shows the frequency of zero-crossing occurrence as a function of wavelength.

length shift variation, supporting the observation that band shift (for this system) is instrumental in nature, as the chemical and physical characteristics of the standard were stable.

When the wavelength shift test was performed using rotating grating monochromator-based instruments, much simpler linear or quadratic wavelength shift error characteristics were observed, which stem from mechanical limitations.²⁵ While it is more difficult to quantify, nonconstant wavelength shift will have a relatively less deleterious effect on predictive performance because the arithmetic mean spectral variation will be lower (unpublished data, Cogdill 2004). However, if linear or quadratic wavelength shift can be accurately quantified, the spectra can be adjusted accordingly. Until more work can be done to fully characterize the wavelength shift variation²⁶ of AOTF instruments, the real-time wavelength accuracy measurement will be used for observation only.

In addition to verification of instrument performance, the calibration monitoring protocol must be able to determine whether or not the calibration models are applicable. Two MSPC tools that are often employed are the lack of model fit, or Q residual statistic, and the sample-to-model distance given by Hotelling's T^2 statistic.⁵ These statistics describe not only model suitability but can detect instrument-induced variation in spectral baseline. For each $I \times p$ sample spectrum, x_i , the Q statistic is calculated as the sum of squared reconstruction error across all wavelengths:

$$Q_{ik} = \sum (x_i - t_{ik} P_k^T)^2, \quad (5)$$

where Q_{ik} is the sum of squared reconstruction error for i^{th} sample spectrum using model k , x_i is the sample spectrum,

t_{ik} is the latent variable scores for i^{th} sample spectrum, using model k , and P_k is the model k loadings, or eigenvectors.

The Hotelling's T^2 statistic is calculated as follows:

$$T_{ik}^2 = t_{ik}(\lambda_k^{-1})t_{ik}^T,$$

where T_{ik}^2 is Hotelling's T^2 for the i^{th} sample spectrum using model k , t_{ik} is the latent variable scores for i^{th} sample spectrum, using model k , and λ_k^{-1} is the diagonal matrix of normalized eigenvalues of the covariance matrix for model k .

The normalized eigenvalue matrix is calculated for model k as follows:

$$\lambda_k = \frac{(T_k^T T_k)}{n_k}, \quad (7)$$

where T_k is the matrix of calibration sample model scores for model k , and n_k is the number of calibration samples used for the estimation of model k .

As seen from Equations 6 and 7, Hotelling's T^2 is a measurement of statistical distance in score space. Hence, T^2 is not influenced by spectral variation out of the plane of the model (Figure 8). A large T^2 indicates that the sample has high leverage on the model and may exceed the confidence limits of the model hyperspace. In these cases, predictions

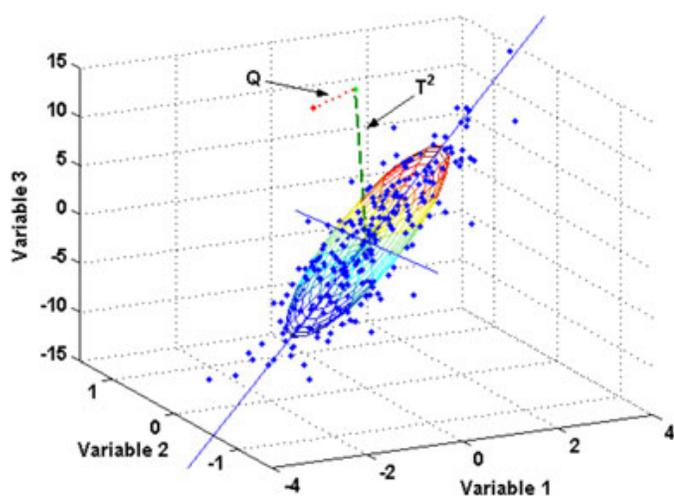


Figure 8. Example of a 2-component PCA model of 3-dimensional data. The primary and secondary component axes are shown scaled by the magnitude of their corresponding eigenvalue. The 2-dimensional ellipse represents the Hotelling's T^2 95% confidence region. The red point illustrates an outlying sample, with its projection on the model (green point). The green dashed line (in the plane) is the distance from the sample projection to the model center (Hotelling's T^2). The red dotted line is the distance from the sample point to the in-plane projection (Q residual).

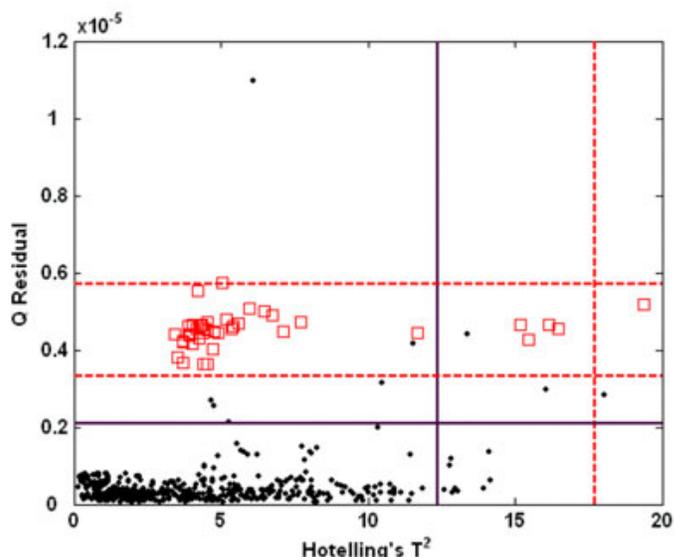


Figure 9. Hotelling's T^2 and Q residual scores for API content calibration (black dots) and VAL2 (red squares) spectra. The solid black lines illustrate T^2 and Q upper control limits estimated using the calibration data. The dashed red lines illustrate upper and lower control limits estimated from the VAL2 scores. The control limits represent 99% confidence intervals.

should be considered invalid. Q , on the other hand, is a measurement of the spectral variation orthogonal to the plane (Figure 8), which is unexplained by the model. Thus, Q and T^2 are completely independent measurements of spectral character. A large Q residual indicates that the sample is poorly reconstructed by the model, which is an indication that either a new factor may be present in the sample matrix or an instrumental fault has occurred (eg, sampling error, component failure).

The Hotelling's T^2 and Q residual statistics for the API content calibration and VAL2 (production-scale) data sets are shown in Figure 9. All of the VAL2 Q residuals, and a portion of the VAL2 T^2 scores, exceed the 99% confidence limits set for the calibration data set. During validation of the API content calibration,² it was discovered that a 6.0-mg bias correction was required for the VAL2 and VAL3 data set predictions. The increase in Q residuals indicates a change in the spectrum shape, which supports the notion that the bias was the result of intercontinental transport and reassembly of the NIR analyzer, as well as normal instrumental drift due to aging (since years had transpired between calibration development and analysis of the VAL2/VAL3 sample sets).² Ideally, a rescue set of stored tablets with known spectra would be recalled to derive a calibration transfer matrix, but no such transfer set was available; thus, the only recourse was either calibration update or correction of the predicted values. Since the observed bias was significant, and attributable to a known event, a single-term bias correction was added to the calibration regression model. As more

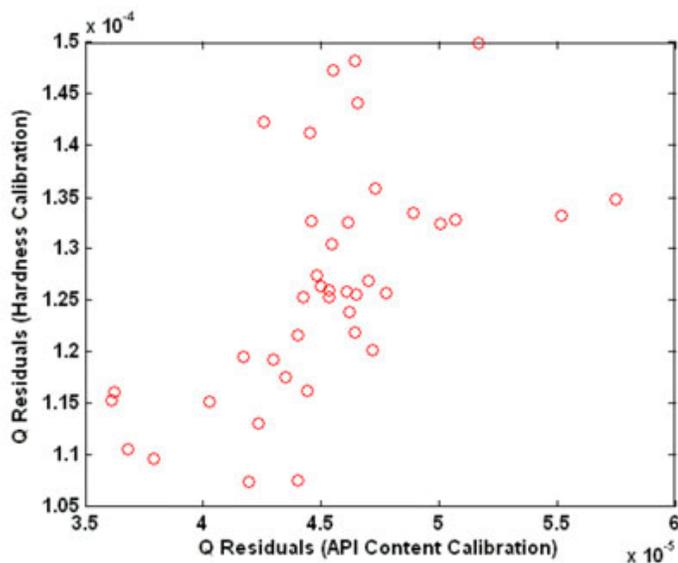


Figure 10. Comparisons of Q residual scores for the VAL2 spectra using the API content and hardness calibration models.

spectral and reference data become available, calibration update and revalidation are the preferred course of action.

Calibration monitoring can also be used as a product identity verification tool. NIR spectra were acquired (using the same instrumentation) for 2 other drug products produced at the same tablet facility. When the API content calibration model was applied to the data, the T^2 and Q scores (not shown) were far higher than was observed for the VAL2 data. The T^2 and Q 99% upper confidence limits for the VAL2 data set were 17.7, and 5.8×10^{-5} , respectively. The mean T^2 scores for the 2 other products were 52.4 and 111.3, and the Q residual scores were 1.4×10^{-4} and 6.4×10^{-4} .

The spectral differences of the VAL2 and false-product data were discernable without need for control limits on the Q and T^2 statistics. Determining control limits on T^2 and Q scores is a complex issue. Control limits (eg, 95% or 99% confidence interval) could be set by assuming statistical distributions for the Q and T^2 scores (Figure 9). This practice provides a high probability that a fraction of otherwise acceptable predictions will be rejected for excessive Q or T^2 scores (eg, 1 tablet per 100 at 99%). This leads to unwarranted shut-down for investigation of the process or analytical method. Rather than setting hard statistical limits a priori, it is suggested that a period of online observation with parallel testing be used to establish control limits. Consistent with this strategy, upper and lower control limits are specified (Figure 9) subsequent to revalidation for prediction of new samples at elevated Q residual values (VAL2). While specification of lower control limits is not a standard practice for these statistics, it should be noted that changes in the Q and T^2 scores (from a set trend), regardless of the direction, indicate a change in

either the process or the analytical method, and should be treated as an excursion from validation.

Since the Q and T^2 statistics will be unique for every calibration model, they can be used as another means of assessing model independence. By plotting T^2 and Q scores from the API content calibration against those from the hardness calibration, it can be seen that their control limits will be different (Figures 10 and 11). The T^2 scores for the 2 calibrations are highly correlated (Figure 11). The significant correlation between the T^2 scores is not unexpected as the spectra and preprocessing for both models were quite similar. Moreover, both models explain greater than 90% of the spectral variance in a common hyperspace. The bias and scale difference between the T^2 scores for the 2 models is due to a difference in model center and degrees of freedom since each calibration model was estimated using different training samples (Equation 6).

The Q residual scores for the 2 calibration models are much less correlated (Figure 10). Since the API content calibration includes an additional latent variable (vs the hardness calibration), a different proportion of spectral variation will be captured, and each model will emphasize a unique set of spectral features. This disparity of emphasis is the root cause for the lack of correlation in Q residual scores, but highlights an advantage of latent variable projection models in specificity. By virtue of projection, the variance that affects each model can be controlled.

Calibration Transfer

FIR^{7,24} is a spectral preprocessing operation where multiplicative scatter correction (MSC)²⁷ is applied recursively as a sliding window operation. As a windowed technique,

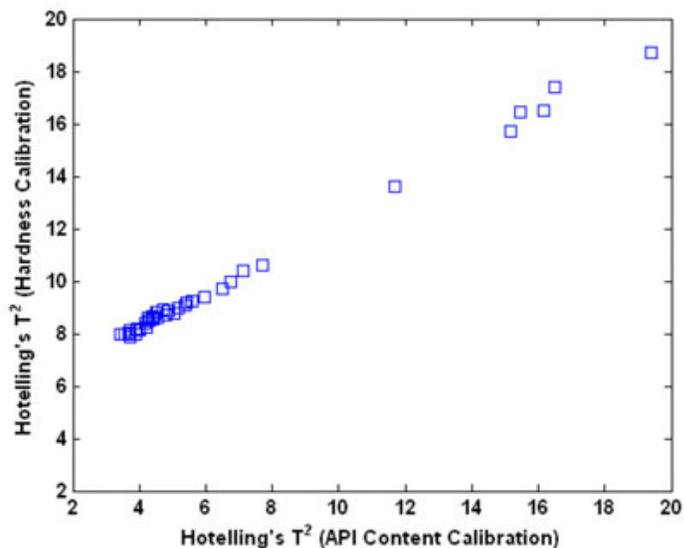


Figure 11. Comparisons of Hotelling's T^2 scores for the VAL2 spectra using the API content and hardness calibration models.

FIR is more flexible than MSC and is able to mitigate complex nonlinear additive and multiplicative baseline effects, whereas MSC adjusts entire spectra with a single set of global slope and offset coefficients. The flexibility of the FIR operation is controlled by adjusting the width of the window, and by changing the shape of the reference spectrum. As window width increases, FIR begins to converge to the MSC solution. As window width decreases, FIR increasingly forces the sample spectra to fit the reference spectrum, eventually removing all spectral variation. As a preprocessing technique, FIR must be performed prior to calibration. Since no model is required, the technique presents a potentially elegant solution for both scatter correction and calibration transfer.

GLS weighting was explored as a DO technique for calibration transfer. GLS is quite similar to the DO routine shown in Equation 1. Rather than subtracting projections of variance, however, GLS down-weights the spectral features to be suppressed. Singular-value decomposition (SVD) is used to create a linear model of the difference spectra. Sample spectra are then normalized by the weighted inverse of their projection on the singular vectors. The GLS weighting is adjusted by scaling the singular values of the projections with an adjustable parameter, a . As a increases (maximum value = 1), the GLS weighting is reduced, and only the greatest singular values have significant influence. As a decreases (minimum value > 0), GLS weighting increases, and increasing emphasis is put on the smaller singular values. GLS must also be performed prior to calibration. Because it is an empirical technique, a set of GLS model coefficients must be maintained. Moreover, for every recalculation of the GLS model, the calibration models must be modified and revalidated.

PDS (described in the Introduction) was applied as an example of a regression-based calibration transfer algorithm. PDS transformation is adjusted by varying window width. With the window width set to 1, univariate regression is performed between corresponding master and slave wavelengths. As the window width is increased, a wider spectral band and latent variable regression are used to model each wavelength of the corrected spectrum. Window widths greater than 1 allow for correction of wavelength shifts and bandpass differences between the master and slave instruments. Furthermore, increasing the window width has an implicit smoothing effect. While regression-based transformation does require storage of coefficients, these algorithms have the advantage of being independent of the calibration. Thus, a data transformation matrix can be modified and stored for each instrument (Figure 1) and can be recalculated and revalidated without adjustment or revalidation of the calibrations.

The calibration transfer training spectra are shown before and after application of the data transformation algorithms

in Figure 2 and Figure 12. For all 3 algorithms, FIR, GLS, and PDS, the master and slave training spectra can be superimposed after transformation. FIR and GLS significantly altered the appearance of both the master and slave spectra. As seen in Figure 12, FIR filtering greatly reduced the amount of variance in the spectra. When the API content calibration was recalculated with FIR and Savitsky-Golay first derivative preprocessing, predictive ability was reduced, suggesting that some pertinent spectral variation was lost as a result of the filtering operation. When applied to the calibration transfer test set, the FIR slave predictions were the least similar to the predictions from the master instrument data (Tables 1 and 2). A large slope and bias were observed for the slave predictions using FIR transformation. The bias was grossly dissimilar between the VAL2 and VAL3 predictions. The applied bias correction was recalculated for FIR and GLS, since those algorithms involve recalculation of the calibration. In both cases, the bias correction was calculated using the same, independent, sample set.²

GLS preserved more spectral variation than FIR (Figure 12), but the shape of the spectrum was significantly modified. Only MSC preprocessing was applied for recalculation of the calibration model after GLS weighting; Savitsky-Golay first derivative was no longer beneficial. As was observed with FIR, calibration performance was reduced somewhat after GLS weighting. The optimal GLS weighting factor, a , was found to be 0.003, which is a very large adjustment. This is likely a result of the significant baseline difference between the master and slave instrument. The RMSET for the GLS calibration was significantly better than was

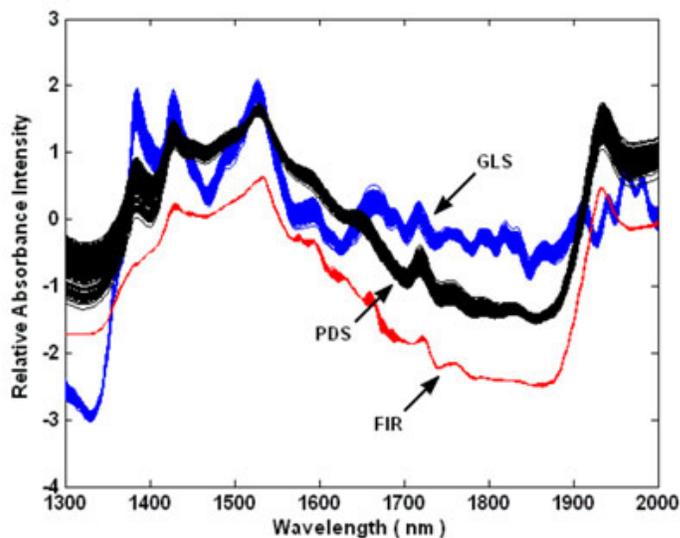


Figure 12. Calibration transfer (slave) training spectra after application of data transformation algorithms (FIR = red, GLS = blue, PDS = black). The spectra were centered and scaled using standard normal variate (SNV) transformation to facilitate presentation.

Table 1. Calibration Transfer Test Results for the VAL2 Samples*

VAL2	Master	FIR	GLS	PDS	Baseline Subtraction
Algorithm parameter	NA	23	0.003	1	NA
RMSET (mg)	NA	4.03	0.98	0.90	1.20
RMSEP (mg) [†]	1.04	2.93	2.97	1.33	1.22
RMSEP (% , nominal) [†]	2.08	5.86	5.94	2.65	2.43
Bias (mg) [†]	0.71	-2.82	-2.52	1.06	0.30

*FIR indicates finite impulse response filtering; GLS, generalized least squares; PDS, piecewise direct standardization; NA, not applicable; RMSET, root mean square error of transfer; and RMSEP, root mean square error of prediction.

[†]Predictions have been corrected for bias using an independent sample set.

observed after FIR filtering. While the VAL3 prediction error for the master instrument would decrease to 3.20 mg after GLS weighting, VAL2 prediction error would increase to 1.44 mg. Though a bias (and slope) difference was observed, it was constant for both the VAL2 and VAL3 samples, suggesting a common bias correction for VAL2 and VAL3. If both the slope and bias were corrected, transfer prediction error (using GLS) could be reduced to 1.33 and 2.91 mg for VAL2 and VAL3, respectively. Furthermore, while bias adjustment can be performed confidently with relatively few samples, slope adjustment of predicted values should be approached with more caution. This, along with the need to constantly recalculate the calibration for every transfer situation, makes GLS weighting a rather complex technique for calibration transfer.

Among FIR, GLS, and PDS, PDS was found to perform the best when applied to production-scale sample spectra (VAL2). As mentioned earlier, PDS has a distinct advantage over the other methods in that it does not involve modification of the calibration or master instrument data, which simplifies the calibration transfer process. When applied to more diverse, laboratory-scale sample spectra, a slope and bias were observed between the master and slave predictions. Since the calibration transfer training spectra were drawn from production-scale tablets, there was rotational ambiguity in the slope estimation due to insufficient covariance between the master and slave responses.

Significant bias is observed when comparing raw absorbance intensity for the master and slave instruments at a single wavelength (Figure 13). However, the difference in slope is small. Recognizing the variability in intensity, and lack thereof in slope, a fourth method of calibration transfer was attempted by simply subtracting the mean spectral difference between the master and slave instruments (Tables 1 and 2). This “baseline subtraction” method²² was observed to perform better than the other methods (Figure 14).

Monte Carlo simulation was used to determine how many calibration transfer samples would be required for successful implementation of the baseline transfer method. During the simulation, varying numbers of training spectra were drawn randomly from the calibration transfer training data set, baseline subtraction was calculated, and the VAL2 prediction error was calculated. As seen in Figure 15, satisfactory calibration transfer is achieved with as few as 15 calibration transfer training samples. Since only additive spectral differences are corrected, production-scale sample spectra are suitable for calibration transfer using baseline subtraction.

The stability sample set calibration monitoring statistics and API content predictions plotted against time of analysis are shown in Figure 16. There is no apparent trend in predicted API content, NFL, wavelength accuracy, or Hotelling's T^2 , but there is a trend of increasing Q residuals with time. Thus, the changes taking place in the samples are in a

Table 2. Calibration Transfer Test Results for the VAL3 Samples*

VAL3	Master	FIR	GLS	PDS	Baseline Subtraction
Algorithm parameter	NA	23	0.003	1	NA
RMSET (mg)	NA	3.77	2.14	4.08	3.25
RMSEP (mg) [†]	3.76	7.23	4.6	5.82	3.99
RMSEP (% , nominal) [†]	7.52	14.46	9.2	11.64	7.98
r^2	0.948	0.979	0.955	0.96	0.927
Bias (mg) [†]	2.04	1.71	-2.41	2.25	-0.84
Slope	1.04	1.98	0.84	1.52	1.08

*FIR indicates finite impulse response filtering; GLS, generalized least squares; PDS, piecewise direct standardization; NA, not applicable; RMSET, root mean square error of transfer; and RMSEP, root mean square error of prediction.

[†]Predictions were corrected for bias using an independent sample set.

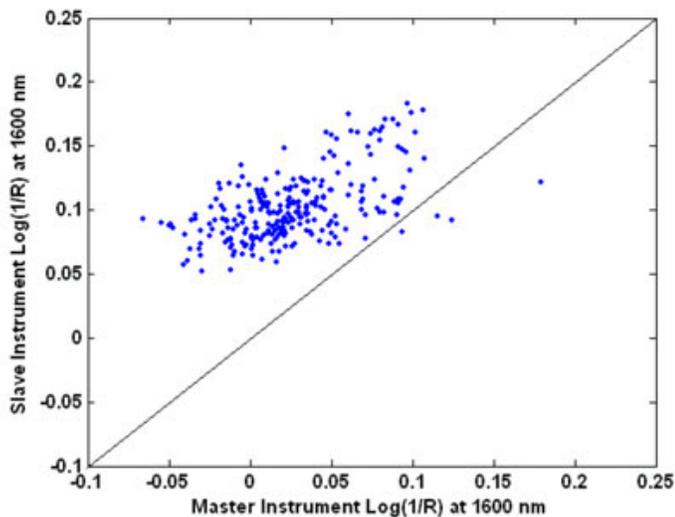


Figure 13. Comparison of NIR absorbance at 1600 nm for the master and slave spectrometers. While little difference in slope is observed, a bias difference is present.

direction orthogonal to the API content prediction model. Depending on what is to be measured, other calibrations might be affected by such changes in tablet quality.

Because the stability test was performed in an uncontrolled environment, and the samples were frequently removed from their sealed containers, it was surmised that the spectral changes were the result of repeated exposure to ambient conditions and handling. The stability test conditions can be considered as extremely adverse. In a more realistic scenario, the rescue samples will be periodically updated and stored under controlled conditions. As shown in Figure 16, except

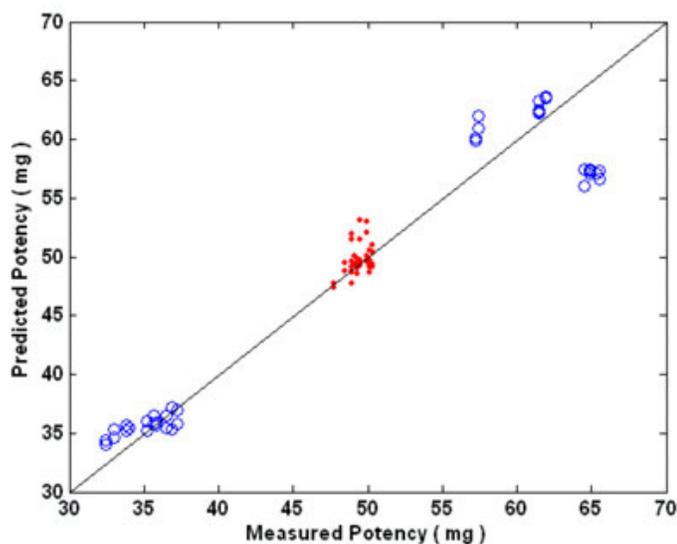


Figure 14. Prediction plot for VAL2 (red dots) and VAL3 (blue circles) slave spectra after calibration transfer using baseline subtraction. Predicted values have been corrected for bias.

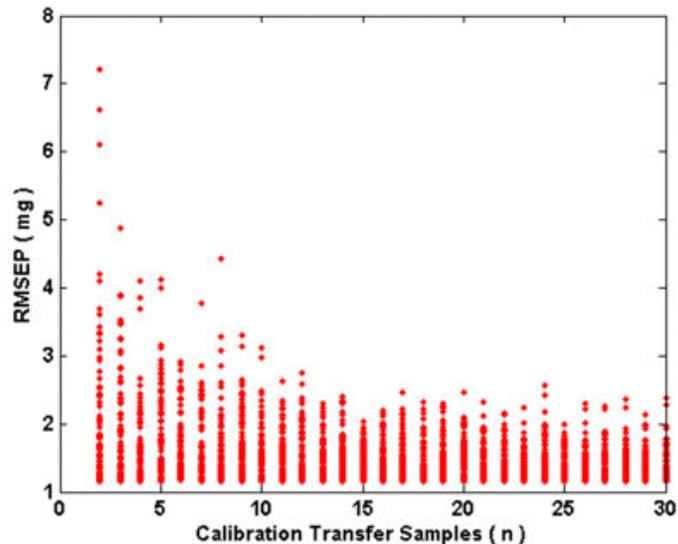


Figure 15. Results of simulated calibration transfer using baseline subtraction with a variable number of calibration transfer samples. Since variation in RMSEP is not greatly reduced after 15 samples, there is little benefit to using a larger calibration transfer data set.

for a single day, calibration transfer using the 9-sample stability set was successful, with VAL2 prediction error below 1.5 mg.

Baseline subtraction calibration transfer was performed to mitigate the spectral effect of lamp change as an example of the use of a calibration transfer rescue set. A set of 16 production-scale tablets were scanned (using the master instrument) 10 times each, with replacement, before and after changing the lamp. Half of the tablets were used for estimation of the baseline subtraction spectrum, and half were used as a test set. A very small, yet complex, difference was observed between the spectra before and

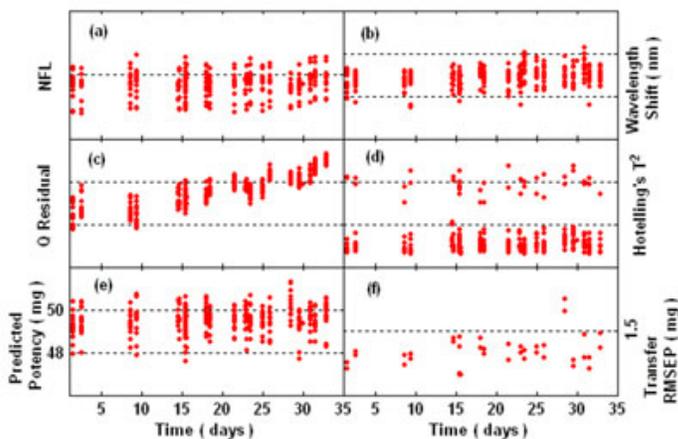


Figure 16. Continuous calibration monitoring scores (A-D), API predictions (E), and RMSEP following calibration transfer (F) for the stability sample spectra as a function of time.

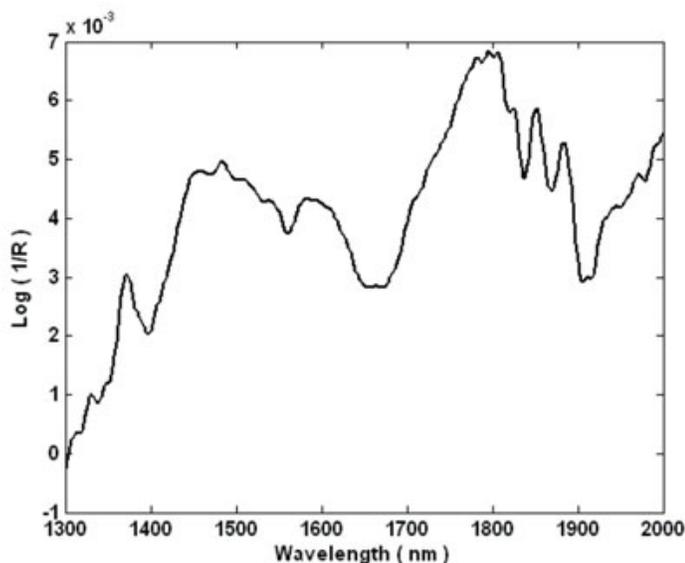


Figure 17. Mean spectral difference between instrument response before and after routine lamp change used for calibration transfer by baseline subtraction. Note that while the curve is complex in shape, the magnitude of the spectral difference is relatively small.

after the lamp change (Figure 17), leading to a -0.32 -mg prediction bias error. After applying the baseline subtraction vector to the test spectra, the prediction bias was reduced to less than 0.05 mg. While the -0.32 mg bias is small enough to be corrected with a simple prediction bias adjustment, long-term instrument maintenance with periodic baseline adjustment may increase the value of spectral databases by reducing a known source of extraneous spectral variation. For this instrument, it is surmised that baseline shape variation is related to polarization dependency of the AOTF crystal, or light source positioning effects.

CONCLUSION

The objectives of this work were the following:

1. Develop a system for continuous calibration monitoring.
2. Formulate an appropriate strategy for calibration data transformation to support instrument maintenance and calibration transfer between instruments.
3. Determine the required number, and allowable storage time, of instrument standardization “rescue” samples.

With these objectives in mind, the experimental results presented for these data demonstrate the following:

1. Key instrument performance parameters can be continuously monitored by analyzing features of sample spectra.

2. Hotelling’s T^2 and spectral Q residuals provide an effective basis for detecting in- and out-of-model spectral deviations.
3. Calibration transfer among multiple instruments can be accomplished with baseline subtraction, and as few as 15 transfer samples for this system.
4. For this application, calibration transfer samples can be stored for at least 1 month without compromising calibration transfer performance.
5. Long-term spectral database uniformity can be maintained using appropriate calibration transfer techniques.

Calibration monitoring provides a means for assessing instrument performance and calibration suitability simultaneously, in real-time. The combination of techniques (NFL, wavelength uncertainty, Hotelling’s T^2 , and Q residual), which are performed for every spectrum, constantly affirm measurement validity.

It is important to note that, while multiple calibration transfer algorithms were compared, the experimental results presented here are specific to this system, and may not reflect general applicability. Rather, the studies performed for this application should be repeated independently for any subsequent analytical application where a calibration must be monitored, maintained, or transferred.

In this, and in prior articles in this series,^{1,2} the development of an online pharmaceutical tablet analysis system using NIR spectroscopy has been illustrated. The deployment and operation of the method will provide opportunities in the future to demonstrate full-scale, real-time calibration monitoring, calibration update, and parallel testing.

ACKNOWLEDGMENTS

Additional data collection was performed by David Molseed, Duquesne University. This work was funded through an agreement between DCPT and AstraZeneca.

REFERENCES

1. Cogdill RP, Anderson CA, Delgado-Lopez M. Process Analytical Technology Case Study, Part I: Feasibility Studies for Quantitative NIR Method Development. *AAPS PharmSciTech*. 2005;6:E262–E272.
2. Cogdill RP, Anderson CA, Delgado-Lopez M. Process Analytical Technology Case Study, Part II: Development and Validation of Quantitative for Tablet API Content and Hardness. *AAPS PharmSciTech*. 2005;6:E273–E283.
3. Food and Drug Administration. *PAT - A Framework for Innovative Manufacturing and Quality Assurance, Draft Guidance*. Rockville, MD: 2003.
4. Box GEP, Jenkins GM, Reinsel G. *Time Series Analysis*. Englewood Cliffs, NJ: Prentice Hall; 1994.
5. Jackson JE, Mudholkar GS. Control procedures for residuals associated with principal components analysis. *Technometrics*. 1979;21:341–349.

6. Williams P, Norris K. *Near-Infrared Technology in the Agricultural and Food Industries*. St. Paul, MN: American Association of Cereal Chemists; 2001.
7. Greensill CV, Wolfs PJ, Speigelman CH, Walsh KB. Calibration transfer between PDA-based spectrometers in the NIR assessment of melon soluble solids content. *J Appl Spectrosc*. 2001;55:647–653.
8. Fearn T. Standardisation and calibration transfer for near infrared instruments: a review. *J Near Infrared Spectrosc*. 2001;9:229–244.
9. Zeaiter M, Roger JM, Bellon-Maurel V, Rutledge DN. Robustness of models developed by multivariate calibration. Part I: the assessment of robustness. *Trends Analyt Chem*. 2004;23:157–170.
10. Fearn T. On orthogonal signal correction. *Chemom Intell Lab Syst*. 2000;50:47–52.
11. Sjöblom J, Svensson O, Josefson M, Kullberg H, Wold S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemom Intell Lab Syst*. 1998;44:229–244.
12. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. *Chemom Intell Lab Syst*. 1998;44:175–185.
13. Andersson CA. Direct orthogonalization. *Chemom Intell Lab Syst*. 1999;47:51–63.
14. Haaland DM, Melgaard DK. New prediction-augmented classical least squares (PACLS) methods: Application to unmodeled interferences. *J Appl Spectrosc*. 2000;54:1303–1312.
15. Wise BM, Martens H, Hoy M. Calibration transfer by generalized least squares. Eigenvector Research Incorporated Report. Available at: <http://www.eigenvector.com/Docs/>. Accessed February 4, 2005.
16. Bouveresse E, Massart D, Dardenne P. Calibration transfer across near-infrared spectrometric instruments using Shenk's algorithm: effects of different standardisation samples. *Anal Chim Acta*. 1994;297:405–416.
17. Dardenne P. Standardisation of near-infrared instruments, influence of the calibration methods and the size of the cloning set. In: Davies AMC, Cho RK, eds. *Near Infrared Spectroscopy: Proceedings of the 10th International Conference*. Chichester, West Sussex, UK: NIR Publications; 2002:23–28.
18. Shenk J. Standardizing NIR instruments. In: Biston R, Bartiaux-Thill N, eds. *Third International Conference on Near-Infrared Spectroscopy*. Gembloux, Belgium: Agricultural Research Centre Publishing; 1991:649–654.
19. Welle R, Greten W, Bernhard R, et al. Near-infrared spectroscopy on chopper to measure maize forage quality parameters online. *Crop Sci*. 2003;43:1407–1413.
20. Wang Y, Veltkamp D, Kowalski BR. Multivariate instrument standardization. *Anal Chem*. 1991;63:2750–2756.
21. Wang Y, Kowalski BR. Temperature-compensating calibration transfer for near-infrared filter instruments. *Anal Chem*. 1993;65:1301–1303.
22. Wang Z, Dean T, Kowalski BR. Additive Background Correction in Multivariate Instrument Standardization. *Anal Chem*. 1995;67:2379–2385.
23. Gallagher NB. Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: improving robustness through model updating. *IFAC ADCHEM '97*. 1997. Available at: www.eigenvector.com/About/NBGCv.html. Accessed February 4, 2005.
24. Wise BM, Ricker NL. Identification of finite impulse response models with principal components regression: frequency-response properties. *Process Contr Qual*. 1992;4:77–86.
25. Funk DB. New methods for wavelength standardisation for near-infrared spectrophotometers, Part 1: review of current standardisation methodology. *J Near Infrared Spectrosc*. 1996;4:101–106.
26. Manning CJ, Griffiths PR. Noise sources in step-scan FT-IR spectrometry. *J Appl Spectrosc*. 1997;51:1092–1101.
27. Martens H, Næs T. *Multivariate Calibration*. New York, NY: John Wiley and Sons; 1989.